Proceedings of the 2005 IEEE
International Conference on Robotics and Automation
Barcelona, Spain, April 2005

A visual odometer without 3D reconstruction for aerial vehicles. Applications to building inspection*

Fernando Caballero, Luis Merino, Joaquín Ferruz and Aníbal Ollero

Departamento de Ingeniería de Sistemas y Automática
Escuela Superior de Ingenieros
41010 Sevilla, Spain
[caba, merino, ferruz, aollero]@cartuja.us.es

Abstract – **This paper presents a vision-based method to estimate the *real motion of a single camera from views of a planar patch. Projective techniques allow to estimate camera motion from pixel space apparent motion without explicit 3-D reconstruction. In addition, the paper will present the HELINSPEC project, the framework where the proposed method has been tested, and will detail some applications in external building inspection that make use of the proposed techniques.**

Index Terms – feature matching, homography, visual odometer, UAV, building inspection.

I. INTRODUCTION

In most cases, unmanned aerial vehicles (UAVs) use GPS to obtain their position. It is well known that the GPS accuracy directly depends on the number of satellites used to estimate the position. This number can be clearly insufficient on urban environments due to buildings. Furthermore, in other scenarios (i.e. forest scenarios), mountains or valleys can also reduce the satellite visibility. The problem may appear when the position of the UAV relies only on the GPS, because usually aerial vehicles lack other backup positioning systems like odometry, which can be easily implemented in terrestrial robots.

It is clear that UAVs need a very robust positioning system because errors in position estimation can generate incoherent control actions and lead to UAV crash and the loss of valuable hardware.

The HELINSPEC project considers the inspection of buildings on urban environments by using UAVs. The main goal is the detection of heat leakage through the building walls and windows, and the search of structure cracks by using fusion techniques between visual and infrared images. As previously stated, a system to estimate the position of the helicopter is needed, because flight near buildings walls sharply reduces satellite visibility and, obviously, increases the risk of accidental crash.

The proposed visual odometer is intended to act as a backup when the accuracy of GPS is reduced to critical levels. The visual odometer avoids explicit 3D reconstruction to reduce the uncertainty of the helicopter position. It will also be shown that the method can also be used in other applications like UAV orientation recovery relative to a plane. The absence of 3D reconstruction allows to reduce the computational load and to increase the number of estimation cycles per second.

The concept of visual odometer [1] was implemented in the CMU autonomous helicopter which demonstrated autonomous visual tracking capabilities of moving objects. Computer vision methods are also used for safe landing in [2]. In [3] a system for helicopter landing on a slow moving target is presented. Also, motion estimation, object identification and geolocation by means of computer vision are described in [4] and [5], in the framework of the COMETS project. Finally, Peter Corke et al. implemented related work using stereo vision for height estimation on a UAV ([6]).

In the following sections the HELINSPEC project will be introduced. After this, a image matching method and homography computation procedure are summarized. Finally, the visual odometer algorithms are explained. And a list of potential applications is presented.

II. THE HELINSPEC PROJECT

The main objective of the HELINSPEC project is to develop new inspection techniques for buildings and other environments with clear accessibility problems, in order to reduce risky and possibly expensive human activity.

These environment make necessary the availability of systems capable of placing and keeping sensors in the desired position in order to acquire all required information for inspection and process all or part of it to get the necessary measurements. In HELINSPEC infrared and visual cameras are being used, as well as distance measurement sensors like laser–based telemeters. On the other hand, the system has to refer the obtained data to maps, so that the real or local coordinates of the inspected object can be known.

Many of the problems found in external building inspection can be solved using aerial platforms with reduced dimensions that allow to place cameras and other sensors close to the inspection objective. It is clear that in many cases the use of aerial vehicles provides significant advantages if compared to other techniques used to deploy and locate the required sensors and instruments.

HELINSPEC uses an autonomous helicopter due to its maneuverability and hover flight capability. A previously existing teleoperated helicopter has been upgraded by providing autonomous flight capabilities and increasing the payload, so that the necessary controller, on-board sensors and communication systems can be safely integrated.

* This research work has been partly supported by the HELINSPEC (PLADIT 2001-2003, SC/AT/0045/2003) and CROMAT (DPI2002-04401-C03-03) projects.

III. Image Matching and Homography Computation

The proposed approach involves three steps: Image matching, homography computation and motion estimation. The image matching step will provide for each pair of consecutive images a sparse set of matching point to point pairs, which will be used in the homography computation. Finally, in the next section, the homography matrix is used to estimate the motion parameters.

A. *Image matching*

The image matching approach used in this application is basically the described in [7], and will not be explained in detail here. A sparse set of corner features is automatically selected in each image; each feature is represented by a fixed-size window which is used as a template. The disambiguation criteria between potential matching pairs combine local similarity obtained by local normalized correlation and cluster-based shape similarity. Clusters are persistent structures which are expected to remain stable for a number of frames, and are searched as a whole. The disambiguation algorithm reduces the computational load by using a cluster-based predictive approach, which involves hypothesis generation, propagation and verification, similar to the one used in [8] for contour matching, and allows to sharply reduce the need of direct search of matching windows and the processing time. This strategy is specially useful if high image to image apparent motion is expected, which leads to the need of large search zones.

In addition, temporary loss of sequences is tolerated through the prediction of the current window position computed with the known position of windows that belong to the same cluster; this feature allows to deal with sporadic occlusion or image noise.

B. *Homography computation*

Assuming that the feature correspondences between two consecutive images are known, it is necessary to compute the best-fitting homography matrix. As is widely known (see [9]), the homography transformation will only be strictly valid for planar scenes or pure rotation motion. The assumption of planar scene is usually acceptable for aerial images of average ground at usual flight altitudes. Furthermore, building walls are predominantly planar, or at least a dominant plane can be extracted from them.

The homography matrix can be computed from point correspondences. At least four of them are needed; in practice an overdetermined system must be solved. For a robust homography computation, the direct least-squares approach cannot be applied. Significant outlier data are present not only because of erroneous matching; independent motion within the scene, as well as non-planar features, will generate non-fitting data which must be discarded. In our approach, the LMedS is used for this purpose ([10]). The following expression is applied:

$$\min\left[\underset{i}{med}\left(r_i^{\,2}\right)\right] \tag{1}$$

where r_i is the Euclidean distance between the feature position in the previous image $\{x_1, y_1\}$ and the position $\{x_2', y_2'\}$ obtained by transforming the position in the current image $\{x_2, y_2\}$ with the homography:

$$r_i = \sqrt{(x_2 - x_2')^2 + (y_2 - y_2')^2} \tag{2}$$

Thus, the estimator must yield the smallest value for the median of squared residual (r_i) computed for the entire data set. It must be solved by searching in the space of possible estimates generated from the data. A randomly chosen subset of data is used to avoid the exhaustive analysis of the solution space.

Finally, once the outliers are detected, a M-Estimator ([10]) will be applied, in order to fit the best homography to the trusted data space.

IV. Multi-view Vision based odometer

In the previous section it has been shown how the homography that relates two views of the same plane can be computed. In this section the approach to recover the real camera motion without 3D reconstruction in the special case of aerial images will be explained.

A. *Theoretical issues*

The homography that relates two images of the same plane taken from different points of view, can be expressed as ([9]):

$$\mathbf{H}_2 = \mathbf{A} \cdot \mathbf{R}_2 \cdot \left(\mathbf{I} - \lambda \mathbf{t}_2 \cdot \mathbf{n}^T\right) \cdot \mathbf{A}^{-1} \tag{3}$$

where $\mathbf{A}$ is the internal calibration matrix of the camera, $\mathbf{t_2}$ is the position of the second point of view in the first camera coordinate frame, $\mathbf{n}$ is the unitary normal vector that defines the plane in the first camera coordinate frame (in the outward direction), λ is the inverse of the distance to the plane of the first camera and $\mathbf{R_2}$ is the rotation that transforms the first camera coordinate frame into the second camera coordinate frame (see Fig. 1).

Reference [11] shows that, if the camera is calibrated (that is, $\mathbf{A}$ is known), it is possible to obtain from $\mathbf{H_2}$ two valid solutions for $\mathbf{R_2}$, $\mathbf{t_2}$, λ and $\mathbf{n}$, up to a scale factor. Furthermore, if a third view is considered (and its homography $\mathbf{H_3}$ respect the first view) it can be obtained a disambiguated solution due to the fact that $\mathbf{n}$ should be unique [12]. Then, the following algorithm can be used:

1. Using the techniques described in section III, the homographies between a first reference frame and consecutive images are computed $\{\mathbf{H_i}, i = 2,\ldots,N\}$.
2. From each $\mathbf{H}_i$, the two valid solutions for $\mathbf{R}_i$, $\mathbf{t}_i$ and $\mathbf{n_i}$ are computed.
3. Given $N > 2$ views and using the fact that $\mathbf{n_i} = \mathbf{n_2}$ the correct solutions accomplish the expression (3) for a small ε.

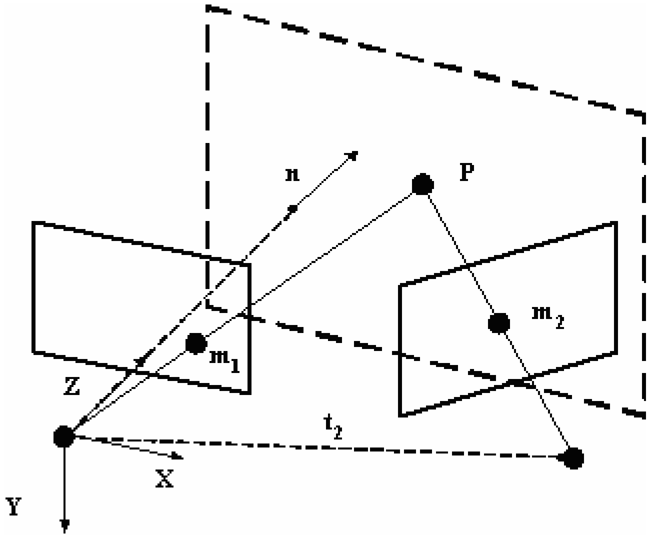


Fig. 1 Geometry between two images of the same plane taken from different points of view.

$$\|\mathbf{n}_i - \mathbf{n}_2\| \leq \varepsilon, i = 3,\ldots,N \qquad (4)$$

Actually, only the product $\mathbf{t}_2 \cdot \mathbf{n}^T$ in (3) can be recovered, so that the translation $\mathbf{t}_2$ is recovered up to a scale factor. If an independent sensor can measure the distance of the first point of view to the plane ($1/\lambda$), this scale factor can be disambiguated. This can be accomplished by means of a sonar or laser range sensor.

B. *Experimental results*

In the following, some of the results that have been obtained under controlled laboratory experiments and field experiments are presented.

In lab experiments a firewire camera connected to a PC computer was used. It was pointed to a quasi-planar scene, and the initial distance to it was measured. In fig. 2 the camera is moved upwards and then to the right for about 17 cm.; then, the movement was reversed in order to return to the first position. It can be seen that the position error at the end of the trajectory is very low, no more that some millimeters.

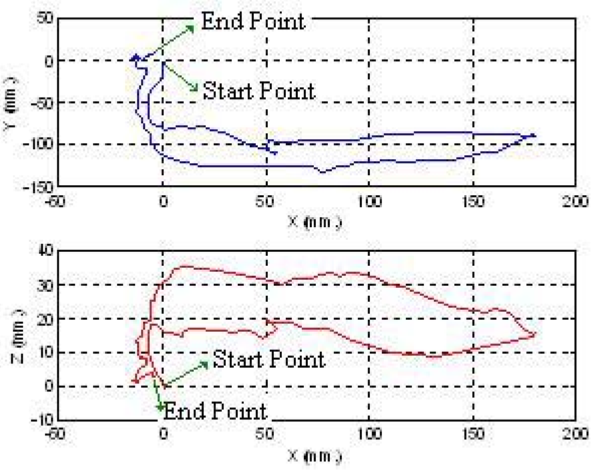


Fig. 2 Firewire camera trajectory in an up and right movement.

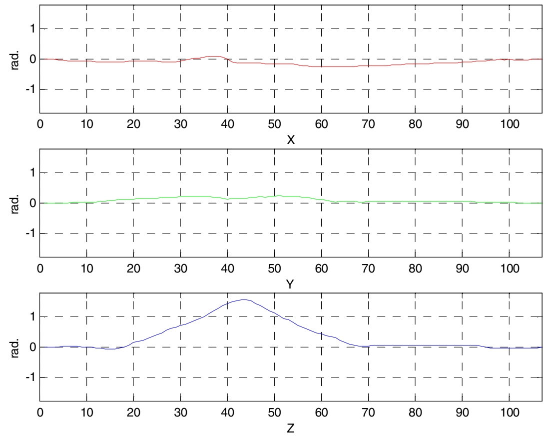


Fig. 3 Firewire camera trajectory in a 90º rotation over Z axis.

In fig. 3 another laboratory experiment is shown; in this case a rotation motion is performed; The camera undergoes a 90º rotation over its Z axis (see figure 1 for the definition of the coordinate frames considered). The figure shows the rotation detected in all axis. It can be seen that the rotation in Z has a peak in 1.5 rad, while the other components show only some residual rotation, because the camera was rotated by hand. Thus, the rotation estimation can be considered correct.

Finally, in Fig. 4, a field experiment is shown. The images were taken by a camera attached to a helicopter, that also carries GPS, IMU and gyros to compute its position and orientation and those of the camera. This experiment was carried out at the Lousa (Portugal) airfield during the general experiments of the COMETS project ([7]). Fig. 5 shows the used COMETS helicopter and fig. 6 shows details about its camera pan&tilt unit.

As it can be seen in Fig. 4, the camera was pointed downwards, but it was not necessarily perpendicular to the ground plane. Notice that the following referenced figures 7, 8 and 9 are represented respect to the number of images of the sequence, to get the time reference only is necessary to know that the image rate is 3 per second.

Fig. 7 details the trajectory of the helicopter and its attached camera in the X,Y and Z axis. GPS measurements (dotted line in fig. 7) are used to check the positions given by the visual odometer system. This figure shows that the errors in X, Y and Z are small, and on the other hand it can be seen that the rate in the estimated position values can be higher than the provided by the GPS.

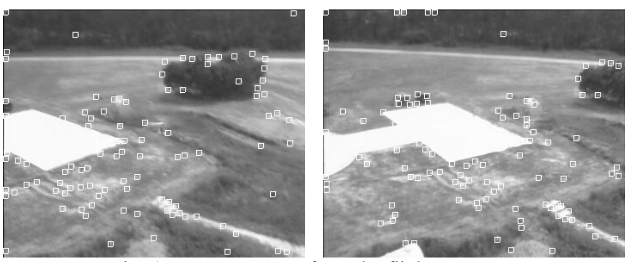


Fig. 4 Image sequence from the flight at Lousa.
White window: Tracked Feature

Fig. 5 COMETS helicopter.

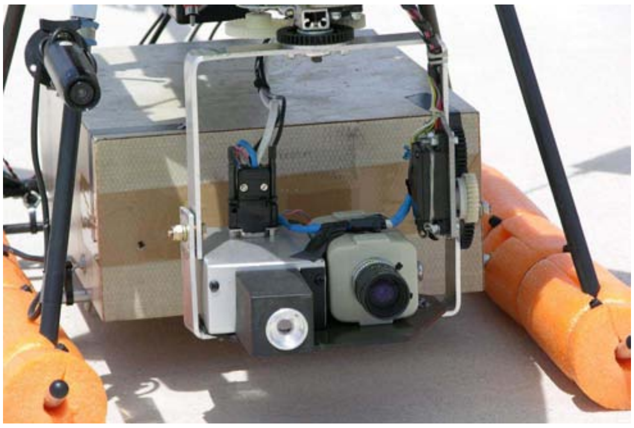


Fig. 6 Camera pan&tilt at COMETS helicopter.

Fig. 8 compares the vision-based estimated orientation with the data offered by the angle sensors of the camera pan&tilt (Fig. 6) and helicopter IMU. The angles are expressed in the first camera reference frame (see Fig. 1), and the evolution of the angles in X,Y and Z camera axis is presented.

Finally, in Fig. 9 the relative errors in the angles estimation are shown. It can be seen that the maximum relative error is around 25%, in X and Z axis, in Y axis the relative error is not upper than 10%

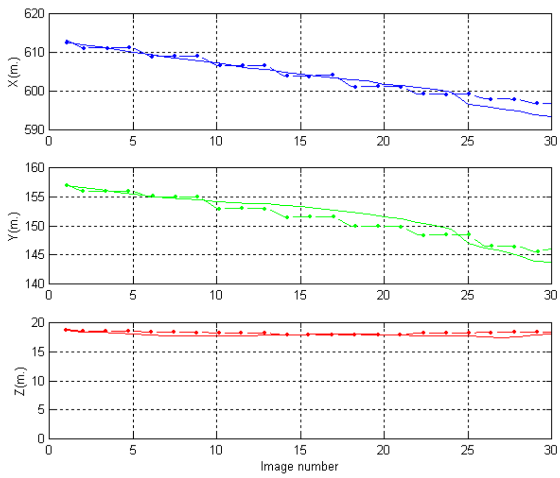


Fig. 7 Flight at Lousa airfield. X,Y,Z coodinates. Dashed: vision-based. Dotted: GPS. 3 images/second

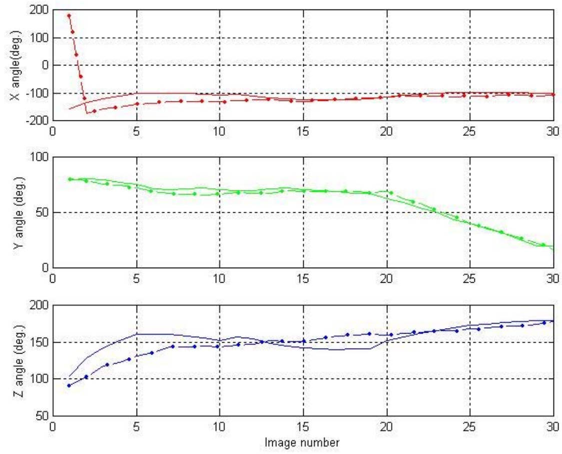


Fig. 8 Flight at Lousa airfield. X,Y,Z angles value evolution. Dashed: vision-based. Dotted: Pan&Tilt sensors. 3 images/second

V. Applications to Building Inspection in the HELISPECT Project

In the framework of the HELINSPEC project, the visual odometer technique is being used not only for UAV positioning when the GPS accuracy is reduced. The proposed technique can estimate the helicopter orientation relative to the building, which is essential for data interpretation and processing, and also to perform approach maneuvers.

A. External building inspection

The visual odometer is able to obtain the relative orientation **n** of the plane respect to the camera. This data can be used to control the helicopter to obtain a perpendicular view of the walls of the inspected building. Also, given an initial estimation of the distance respect to this plane, it is possible to obtain the distance respect to the plane each time a new image is received. Fig. 12 shows the estimated distance respect to the wall using an image sequence at 3 images per second rate.

Moreover, the knowledge of **n** can be used to generate virtual frontal views of the inspected building. The frontal view can be an important tool when it is necessary to recognize geometric features because variations in the camera orientation may create significant distortion.

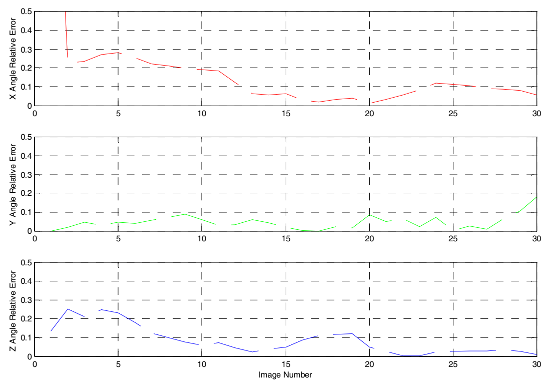


Fig. 9 Flight at Lousa airfield. X,Y,Z angles relative error. 3 images/second

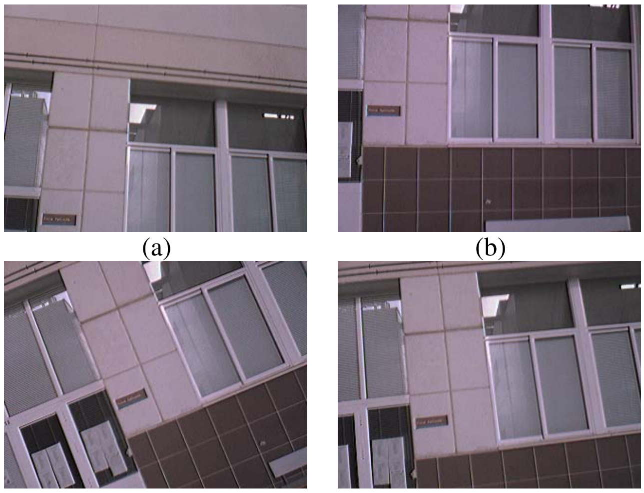
To obtain a frontal view, an approach similar to de presented in [13] can be adopted. Given $\mathbf{n}$, it is needed to apply the rotation $\mathbf{R}$ that makes $\mathbf{n}$ collinear with the direction of the optical axis of the camera (Z axis, see Fig. 10). That is, $\mathbf{R}$ describes a rotation of θ over the axis given by the vector $\boldsymbol{\omega}$, where $\boldsymbol{\omega}$ and θ are computed following (5):

$$\boldsymbol{\omega} = \mathbf{n} \wedge \mathbf{z}$$
$$\theta = \cos^{-1}(\mathbf{n}^T \cdot \mathbf{z}) \qquad (5)$$

The rotation $\mathbf{R}$ can be computed as:

$$\mathbf{R} = (1-\cos(\theta))\boldsymbol{\omega}\cdot\boldsymbol{\omega}^T + \sin(\theta)[\boldsymbol{\omega}]_x + \cos(\theta)\mathbf{I} \qquad (6)$$

where $[\boldsymbol{\omega}]_x$ is the antisymmetric matrix such that for any vector $\mathbf{v}$, $[\boldsymbol{\omega}]_x \mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{v}$. Once $\mathbf{R}$ is known, (3) can be used to compute the homography to synthesize the frontal view.


(a) (b) (c) (d)
Fig. 11 Images 1 (a), 11 (b), 41 (c) and 80 (d) of the sequence.

B. *Experimental results*

In order to test the presented technique under realistic conditions, an image sequence of a building wall was recorded. As shown in Fig. 11, the camera orientations were not perpendicular to the wall plane; the initial distance to the wall was known (5 meters). The camera was moved in an straight line in front of the wall, keeping the distance to the to it approximately constant but modifying the orientation of the camera. The goal was to test the distance computation results of the algorithm.

The results of the experiment can be shown in the Fig. 12 where the distance to the wall can be seen. It is clear that the distance is approximately constant, corresponding to the experiment constraint. In addition, in Fig. 13 the frontal view generated using the non-perpendicular views of the wall is shown. It was necessary to warp the images using the information about the orientation of the wall to obtain the final result presented in this figure. As a comparison, it is also presented the resultant mosaic without the perpendicular rectification

VI. Conclusions and Future Developments

This paper presents a method for UAV position estimation based on image processing which can be considered as a visual odometer. It has been developed in the framework of the HELINSPEC project, whose target application is the visual inspection of building walls. The paper includes laboratory and field experimental results that show the feasibility of the approach.

The visual odometer can be used as an effective backup of GPS. As it can provide position data relative to walls, it will also be useful for key application-dependent functions, like ortophoto and mosaic generation, and approach helicopter maneuvers.

Future developments could include a vision-based landing application similar to the proposed in [12], although in this case no assumption of a known pattern would be necessary; natural landmarks could be used.

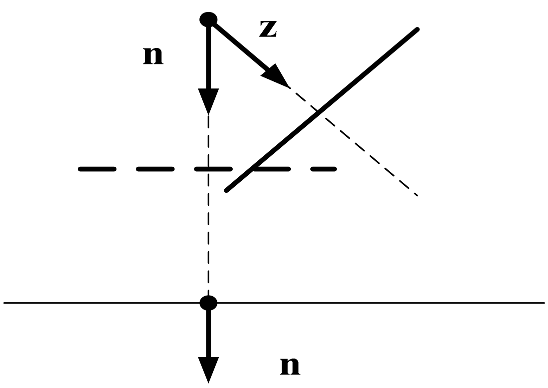


Fig. 10 Orientation estimation in building inspection.

Acknowledgment

The authors thank the cooperation of other members of the COMETS and HELINSPEC teams at Seville.

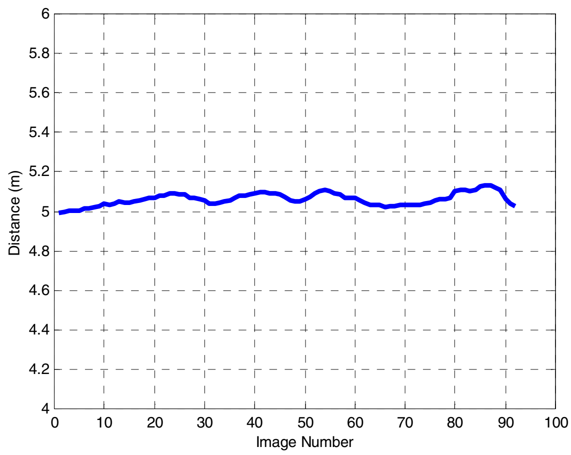


Fig.12 Estimated distance from the wall for the images of figure 11. 3 images/second

Fig.13 : Up: Mosaic using as reference the first frame of the sequence of figure 11. Bottom: wall frontal view using the computed wall normal vector.

REFERENCES

[1] O. Amidi, T. Kanade and K. Fujita, "A visual odometer for autonomous helicopter flight," Proceedings of IAS-5, 1998.

[2] R. Vidal, S. Sastry, J. Kim, O. Shakernia and D. Shim, "The Berkeley Aerial Robot Project (BEAR)," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 1-10. Lausanne, Switzerland, 2002.

[3] P.J. Garcia-Pardo, G.S. Sukhatme and J.F. Montgomery, "Towards Vision-Based Safe Landing for an Autonomous Helicopter," Robotics and Autonomous Systems, Vol. 38, No. 1, pp. 19-29, 2001.

[4] L. Merino and A. Ollero, "Forest fire perception using aerial images in the COMETS Project," 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems – IROS 2002. Proc. Workshop WS6 Aerial Robotics, pp 11-22. Lausanne, Switzerland, 2002.

[5] L. Merino and A. Ollero, "Computer vision techniques for fire monitoring using aerial images," Proc. of the IEEE Conference on Industrial Electronics, Control and Instrumentation IECON 02 Seville (Spain), 2002.

[6] P. Corke, P.Sikka and J. Roberts, "Height estimation for an autonomous helicopter," International Symposium on Experimental Robotics, 2000.

[7] A. Ollero, J. Ferruz, F. Caballero, S. Hurtado and L. Merino. "Motion compensation and object detection for autonomous helicopter visual navigation in the COMETS system". Proceedings of the 2004 IEEE Int. Conf. on Robotics and Automation, pp 19-24, New Orleans, 2004

[8] N. Ayache, "Artificial Vision for Mobile Robots," The MIT Press, Cambridge, Massachusetts, 1991.

[9] O. Faugeras, Q. Luong and T. Papadopoulo, "The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene andSome of Their Applications", MIT Press, 2001.

[10] Z. Zhang, "Parameters estimation techniques. A tutorial with application to conic fitting," RR-2676, Inria, October 1995.

[11] R.Y. Tsai, T.S. Huang and W.-L. Zhu. "Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition" IEEE Trans. Acoustic, Speech and Signal Processing, Vol. 30, no. 4 pp. 525-524, August 1982.

[12] O. Shakernia, R. Vidal, C. S. Sharp, Yi Ma, S. Sastry. "Multiple view motion estimation and control for landing an unmanned aerial vehicle". Proceedings of the 2002 IEEE Int. Conf. on Robotics and Automation, pp 2793-2798, Washington DC, May 11-15, 2002

[13] L. Lee, R. Romano and G. Stein. "Monitoring activities from multiple video streams: establishing a common coordinate frame" IEEE Transactions on PAMI. Vol .22, no. 8, pp. 758-767. August 2000